ERIC Document Resume	See Instruction Sheet	DATE OF RESUME	DATE RECEIVED

A. PUBLICATION DATA

ED001649	2. ORIGINATOR CONTROL NO(S). N.A.	3. IDENTIFICATION NO. N.A.	DO NOT WRITE IN SPACE BELOW

4. AUTHOR(S):
Fishman, Joshua et al.

5. TITLE OF DOCUMENT ABSTRACTED:
Guidelines for Testing Minority Group Children

6. PUBLICATION TITLE:
N.A.

7. EDITOR(S):
N.A.

8. PLACE OF ORIGIN
ORG. Society for the Psychological Study of Social Issues
ST. American Psychological Association
CITY
TEXT
INST/LOC CODE

9. PUBLISHER OR SOURCE
Society for the Psychological Study of Social Issue, American Psychological Assoc.

10. VOL.	11. PAGES	12. MO./YR.	13. TOTAL PP.
N.A.	N.A.	6/63	27

MICROFILM COPY AVAILABLE YES ☒ NO ☐

COPYRIGHT GRANTED YES ☐ NO ☒

COPYRIGHT RELEASE FILED YES ☐ NO ☐

TYPE OF REPORT
PROPOSAL ☒
IN-PROGRESS ☐
FINAL ☐
FOLLOW-UP ☐

OTHER ERIC REFERENCES:

ERIC SATELLITE CODE

B. INDEXING DATA

14. SUBJECT AREA

15. TYPE OF ENDEAVOR
guidelines for testing

16. INDEXING TERMS (6-12)

educational testing	reliability	culture free tests	test interpretation	
psychological testing	test performance	cultural factors	response sets	
minority group	predictive validity	failure barriers	differentiation	
		testing		

17. ABSTRACT (MAX. 250 WORDS)

Educators possess special service and instructional skills which, if used wisely, can assist minority group children in overcoming their early disadvantages. Educational and psychological tests may help if they are carefully and intelligently employed. Professional training and diagnostic sensitivity are required.

Standardized tests currently in use present three principal difficulties when used with disadvantaged minority groups. First, they may not provide reliable differentiation in the range of the minority group scores. Also, many characteristics of minority group children affect test performance. The lower-class child will tend to be less verbal, less self-confident, less motivated toward academic achievement, less competitive intellectually, less exposed to stimulating materials in the home, less knowledgeable about the world, and more fearful of strangers than the middle class child.

The second difficulty is that the significance of the tests for predictive purposes may be quite different for the minority groups than for the standardization and validation groups. Factors that affect the test scores but which may have little relation to the criterion, such as test-taking skills, anxiety and motivation, may impair predictive validity. Also, the criteria which a test is predicting are usually more complex than the test itself. It is important to recognize the influence of such other factors as personality and background, which may be related to criterion performance. Test results cannot reveal the degree to which the status of disadvantaged children might change if environmental opportunities and incentives for learning were improved. Guidance and special training are therefore very important. Tests labeled "culturally unfair" may be valid predictors for school criteria which may be socially unfair. Most "culture free" tests have low predictive validity for academic work.

ERIC No.

18. SOURCE OF ABSTRACT

C. FUNDING DATA

19. SUPPORTING ORGANIZATION	20. DIVISION OR DEPT.	21. CONTRACT NO. / PROJECT NO.
	22. TOTAL COST OF PROJECT	
	23. DURATION OR PROJECT	U.S. Department of Health, Education, and Welfare Office of Education

C Form 1

ED001649

(L-41 L23, L-
L34 T34
(not in L34)
Edmund W. Gordon

INFORMATION RETRIEVAL CENTER ON THE DISADVANTAGED
Ferkauf Graduate School of Education, Yeshiva University

GUIDELINES FOR TESTING MINORITY GROUP CHILDREN

Prepared by a Work Group

of the

Society for the Psychological Study of Social Issues

(Division 9 of the American Psychological Association)

Martin Deutsch
Joshua Fishman, Chairman
Leonard Kogan
Robert North
Martin Whiteman

June 5, 1963

REPRINTED BY PERMISSION
U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN PRINTED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

GUIDELINES FOR TESTING MINORITY GROUP CHILDREN

INTRODUCTION

American educators have long recognized that they can best guide the development of intellect and character of the children in their charge if they take the time to understand these children thoroughly and sympathetically. This is particularly true with respect to the socially and culturally disadvantaged child. Educators must realize that they possess special service and instructional skills of considerable responsibility and power. These, if used wisely, can assist minority group children in overcoming their early disadvantages, fully or at least in part, and can enable them to live more constructively and to contribute more fully to American society. Educational and psychological tests may help or hinder the attainment of this desired state of affairs, depending on how carefully and intelligently the tests are employed. Persons who have a genuine commitment to democratic processes and to the deep respect for the individual, without which no democracy can function, will certainly seek to use educational and psychological tests with minority group children in ways that will enable these children to attain the full promise that America holds out to all its children.

Educational and psychological tests are among the most widely used and useful tools of teachers, educational supervisors, school administrators, guidance workers, and counselors. As is the case with all specialized tools, however, professional training and diagnostic sensitivity are required for the intelligent and responsible use of these instruments. That is why most colleges and universities offer courses in educational and psychological testing for educational personnel. That is also why there are now a number of relatively non-technical books and brochures which seek to acquaint ed-

ucators and their associates with the principles and procedures of proper test selection, use and interpretation.[1]

Responsible educational authorities recognize that it is as dangerous to put tests in the hands of untrained and unskilled personnel as it is to permit the automobile or any highly technical and powerful tool to be handled by individuals who are untrained in its use and unaware of the damage that it can cause if improperly used.

The necessity for caution is doubly merited when educational and psychological tests are administered to members of minority groups. Unfortunately, there is no single and readily available reference source to which test users can turn in order to become more fully acquainted with the requirements and cautions to be observed in such cases. The purpose of this committee's effort is to provide an introduction to the many considerations germane to selection, use and interpretation of educational and psychological tests with minority group children as well as to refer educators and their associates to other more technical discussions of various aspects of the same topic.

The term "minority group" as we are using it here is not primarily a numerical designation. Rather it is a status designation referring to cultural or social disadvantage. Since many Negro, Indian, lower class white, and immigrant children have not had most of the usual mid-class opportunities to grow up in home, neighborhood, and school environment that might enable them better to utilize their ability and personality potentials, they are at a disadvantage in school, and in after-school and out-of-school situations as well. It is because of these disadvantages - vis a vis the majority of American children - disadvantages produced by environmental deprivations and experiential atypicalities - that certain children may be referred to as minority group children.

The following discussion is an elaboration of the technical recommendations developed for various kinds of tests by committees of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education (1954, 1955). Our contribution is directed toward specifying the particular considerations that must be kept in mind when professional educators and those who work with them utilize educational and psychological tests with minority group children.

CRITICAL ISSUES IN TESTING MINORITY GROUPS

Standardized tests currently in use present three principal difficulties when they are used with disadvantaged minority groups: (1) they may not provide reliable differentiation in the range of the minority group's scores, (2) their significance for predictive purposes may be quite different for the minority groups than for the standardization and validation groups and (3) the validity of their interpretation is strongly dependent upon an adequate understanding of the social and cultural background of the group in question.

I. RELIABILITY OF DIFFERENTIATION

In the literature of educational and psychological testing, relatively little attention has been given to the possible dependence of test reliability upon subcultural differences. Although it is considered essential for a test publisher to describe the reliability sample (the reference group upon which reliability statements are based) in terms of factors such as age, sex, and grade level composition, it is more useful if the reliability sample is divided into subgroups in terms of these factors, and separate reliability coefficients reported for each subgroup. There is a growing tendency on the part of test publishers to report subgroup reliabilities

for age, sex, and grade level, but, to the best of our knowledge, none of the test manuals for the widely used tests provide separate reliability data for specific minority groups. Institutions that use tests regularly and routinely for particular minority groups would do well to carry out their own reliability studies in order to determine whether the tests are reliable enough when used with these groups.

Reliability Affected by Spread of Scores and Average Score of Group

In addition to being dependent on test length and the specific procedure used for estimating reliability, e.g., split-half or retest, the reliability coefficient for a particular test is strongly affected by the spread of test scores in the group for which the reliability is established. In general, the greater the spread of scores in the reliability sample, the higher the reliability coefficient. Consequently, if the tester attempts to make discriminations within a group which is more homogeneous than the reference or norm group for which reliability is reported, the actual effectiveness of the test will tend to be lower than the reported reliability coefficient appears to promise. For many tests, there is abundant evidence that children from the lower socio-economic levels commonly associated with minority group status tend to have a smaller spread of scores than do children from middle-income families, and such restriction in the distribution of scores tends to lower reliability so far as discrimination of measurement within such groups is concerned.[2]

Similarly, many tests show different degrees of reliability, depending upon the average score of the group being tested. In this connection, however, there is no clear-cut tendency for the standard error of measurement[3] to decrease as the average level of ability of the group becomes lower. However, for some tests (such as the Stanford-Binet) it is generally reported that there is a decrease in the magnitude of the standard error of

measurement as the IQ's get lower. This does not necessarily indicate that the test provides more precise differentiation at the lower ability levels, indeed it may provide relatively less differentiation when the range of scores is taken into account. On a comparative basis for example, if the IQ's of a group of underprivileged children range from 80 to 100, a standard error of measurement of five points represents a greater proportion of error variance for that group than does a standard error of measurement of six points for a college preparatory group whose IQ's range from 100 to 160.

Characteristics of Minority Group Children that Affect Test Performance

Most of the evidence relating to the contention that the majority of educational and psychological tests tend to be more unreliable, i.e., more characterized by what is technically called "error variance," for minority group children, is indirect, being based on studies of social class and socio-economic differences rather than on minority group performance per se. Nevertheless, the particular kinds of minority groups that we have in mind are closely associated with the lower levels of socio-economic status. The results of studies by Warner, Davis, Deutsch, Deutsch and Brown, Havighurst, Hollingshead, Sears, Maccoby, and many others are cases in point. Many of these studies are discussed by Anastasi (1958), Tyler (1956) and Deutsch (1960).

For children who come from lower socio-economic levels, what characteristics may be expected to affect test performance in general, and the accuracy or precision of test results in particular? The list of reported characteristics is long and it is not always consistent from one investigation to another. But, at least, it may be hypothesized that in contrast to the middle-class child the lower-class child will tend to be less verbal, more fearful of strangers, less self-confident, less motivated toward scholastic and academic achievement, less competitive in the intellectual realm, more

irritable, less conforming to middle-class norms of behavior and conduct, more apt to be bilingual, less exposed to intellectually stimulating materials in the home, less varied in recreational outlets, less knowledgeable about the world outside his immediate neighborhood, and more likely to attend inferior schools.

Some Examples

Can it be doubted that such characteristics - even if only some of them apply to each "deprived" minority group - will indeed be reflected in test-taking and test performance? Obviously, the primary effect will be shown in terms of test validity for such children, but in many cases the lowering of test validity may be indirectly a result of lowered test reliability particularly if such characteristics interfere with the consistency of performance. Consider the following examples and probable results.

Example: A Negro child has had little contact with white adults other than as distant and punitive authority figures. Probable Result: Such a child might have difficulty in gaining rapport with a white examiner or reacting without emotional upset to his close presence. Even in an individual testing situation, he might not respond other than with monosyllables, failing to give adequate answers even when they know them. The examiner, reacting in terms of his own stereotypes, might add to the unreliability of the test results by assuming that the child's performance will naturally be inferior.

Example: Children from a particular minority group are given little reason to believe that doing well in the school situation will affect their chances for attaining better jobs and higher income later in life. Probable Result: Such children will see little purpose in schooling, dislike school, and will reject anything associated with school. In taking tests, their primary objective is to get through as rapidly as possible and escape from what

for them might be an uncomfortable situation. Their test performance might, therefore, be characterized by a much greater amount of guessing, skipping, and random response than is shown by the middle-class child who never doubts the importance of the test, wants to please his teacher and parents, and tries his best.

Special Norms Often Needed

When the national norms do not provide adequate differentiation at the lower end of the aptitude or ability scale, special norms, established locally, are often useful. For instance, if a substantial number of underprivileged or foreign-background pupils in a school or school district rank in the lowest five per cent on the national norms, local norms might serve to provide a special scale within this range. If the test does not actually yield reliable scores within this range, however, (i.e., if the score distribution with the first few percentiles of the national norms is mainly a function of chance factors) a lower level of the test or an easier type of test is needed for accurate measurement of the low-scoring children.

Responsibilities of Test Users

The sensitive examiner should be alert to reliability considerations in regard to the particular groups he is testing and the intended uses of the tests. In assessing reports on test reliability provided by test manuals and other sources, he will not be satisfied with high reliability coefficients alone. He will consider not only the size of the reliability samples, but also the nature and composition of the samples and the procedures used to estimate reliability. He will try to determine whether the standard error of measurement varies with score levels, and whether his testing conditions are similar to those of the reliability samples. He will ask whether the evidence on reliability is relevant to the persons and purposes with which he is concerned. He will know that high reliability does

not guarantee validity of the measures for the purpose in hand, but he will realize that low reliability may destroy validity.

The examiner should be well aware that test results are characteristically influenced by cultural and subcultural differentials and that the performance of under-privileged minority group children is often handicapped by what should be test-extraneous preconditions and response patterns. He should not necessarily assume that the child from a minority group family will be as test-sophisticated and motivated to do his best as are the majority of environment-rich middle class children.

If the examiner finds - and this will be typical - that the reliability sample does not provide him with information about the reliability of the test for the kind of children he is testing, he should urge that the test results not be taken at face value or from the exclusive base for critical decision-making about the children concerned. Very often careful examination of responses to individual test items will indicate to him that the apparent performance of the child is not adequately reflecting the child's actual competence or personality, but rather that the results are largely attributable to subcultural group factors.

II. PREDICTIVE VALIDITY

Of course, if an individual's test scores were only to be used to describe his relative standing with respect to a specified norm group, the fact that the individual had a minority-group background would not be important. It is when an explanation of his standing is attempted or when long-range predictions enter the picture, as they usually do, that background factors become important.

For example, no inequity is necessarily involved if a culturally disadvantaged child is simply reported to have an IQ of 84 and a percentile

rank of 16 on the national norms for a certain intelligence test. However, if this is interpreted as meaning that the child ranks or will rank no higher in learning ability than does a middle-class, native-born American child of the same IQ, the interpretation might well be erroneous.

Factors Impairing Test Validity

Three kinds of factors may impair a test's predictive validity. First, there are test-related factors - factors or conditions that affect the test scores but which may have relatively little relation to the criterion. Such factors may include test-taking skills, anxiety, motivation, speed, understanding of test instructions, degree of item or format novelty, examiner-examinee rapport, and other general or specific abilities that underlie test performance but which are irrelevant to the criterion. Examples of the operation of such factors are found in the literature describing the problems of white examiners testing Negro children (Dreger and Miller, 1960), of American Indian children taking unfamiliar, timed tests (Klineberg, 1935), and of children of certain disadvantaged groups being exposed for the first time to test-taking procedures (Haggard, 1954). It should be noted that some test-related factors may not be prejudicial to disadvantaged groups. For example, test-taking anxiety of a disruptive nature (Sarason et al, 1960) may be more prevalent in some middle-class groups than in lower-class groups. In general, however, the bias attributable to test-related factors accrues to the detriment of the culturally disadvantaged groups.

The problem of making valid predictions for minority group children is faced by the Boys' Club of New York in its Educational Program[4], which is designed to give promising boys from tenement districts opportunities to overcome their environmental handicaps through scholarships to outstanding schools and colleges. Although the majority of the boys currently enrolled in this program had mediocre aptitude and achievement test scores

up to the time they were given scholarships, practically all of the boys had achieved creditable academic success at challenging secondary boarding schools and colleges. In this program, normative scores on the Otis Quick-Scoring Mental Ability Test and the Stanford Achievement Test are used for screening purposes, but they are regarded as minimal estimates of the boys' abilities. The Wechsler Intelligence Scale for Children (WISC) is frequently used in this program to supplement the group tests. The boys typically score 5 to 10 points higher on the WISC than on the Otis, probably because the WISC gives less weight to educational and language factors.

Interest and Personality Inventory Scores

When standardized interest inventories are used, special caution should be observed in making normative interpretations of the scores of culturally disadvantaged individuals. When a child has not had opportunities to gain satisfaction or rewards from certain pursuits, he is not likely to manifest interest in these areas. For example, adolescent children in a particular slum neighborhood might rank consistently low in scientific, literary, musical, and artistic interests on the Kuder Preference Record if their home and school environments fail to stimulate them in these areas. With improved cultural opportunities, these children might rapidly develop interests in vocations or avocations related to these areas.

Scores on personality inventories may also have very different significance for minority group members than for the population in general (Auld, 1952). Whenever the inventory items tap areas such as home or social adjustment, motivation, religious beliefs, or social customs, the appropriateness of the national norms for minority groups should be questioned. Local norms for the various minority groups involved might again be very much in order here.

Predicting Complex Criteria

A second class of factors contributing to low predictive validity is associated with the complexity of criteria. Criteria generally represent "real life" indices of adjustment or achievement and therefore they commonly sample more complex and more variegated behaviors than do the tests. An obvious example is the criterion of school grades. Grades are likely to reflect motivation, classroom behavior, personal appearance, and study habits, as well as intelligence and achievement. Even if a test measured scholastic aptitude sensitively and accurately, its validity for predicting school marks would be attenuated because of the contribution of many other factors to the criterion. It is important, therefore, to recognize the influence of other factors, not measured by the tests, which may contribute to criterion success. Since disadvantaged groups tend to fare poorly on ability and achievement tests (Anastasi, 1958; Tyler, 1956; Masland, Sarason, and Gladwin, 1958; Eels et al, 1951; Haggard, 1954), there is particular merit in exploring the background, personality, and motivation of members of such groups for compensatory factors, untapped by the tests, which may be related to criterion performance.

In some instances, such as in making scholarship awards on a statewide or national basis, test scores are used rigidly for screening or cut-off purposes to satisfy demands for objectivity and "impartiality." The culturally disadvantaged child (quite possibly a "diamond-in-the-rough") is often the victim of this automatic and autocratic system. Recourse lies in providing opportunities where the hurdles are less standardized and where a more individualized evaluation of his qualifications for meeting the criterion may prove to be fairer for him.

For example, the following characteristics that may be typical of minority group children who have above-average ability or talent are among

those cited by DeHaan and Kough (1956), who have been working with the North Central Association Project on Guidance and Motivation of Superior and Talented Secondary School Students:

> They learn rapidly, but not necessarily those lessons assigned in school.
>
> They reason soundly, think clearly, recognize relationships, comprehend meanings, and may or may not come to conclusions expected by the teacher.
>
> They are able to influence others to work toward desirable or undesirable goals.

Effects of Intervening Events on Predictions

A third set of contributors to low criterion validity derives from the nature of intervening events and contingencies. This class of conditions is particularly important when the criterion measure is obtained considerably later than the testing - when predictive rather than concurrent validity is at stake. If the time interval between the test administration and the criterial assessment is lengthy, a host of situational, motivational, and maturational changes may occur in the interim. An illness, an inspiring teacher, a shift in aspiration level or in direction of interest, remedial training, an economic misfortune, an emotional crisis, a growth spurt or retrogression in the abilities sampled by the test - any of these changes intervening between the testing and the point or points of criterion assessment may decrease the predictive power of the test.

One of the more consistent findings in research with disadvantaged children is the decline in academic aptitude and achievement test score of such children with time (Masland, Sarason, and Gladwin, 1958). The decline is, of course, in relation to the performance of advantaged groups or of the general population. It is plausible to assume that this decline represents the cumulative effects of diminished opportunities and decreasing motivation for acquiring academic knowledge and skills. When such cumulative effects

are not taken into consideration, the predictive power of academic aptitude and achievement tests is impaired. On the other hand, if it were known that certain individuals or groups would be exposed to deleterious environmental conditions, allowance might be made for such contingencies in the predictive formulas and the test's criterion validity could be improved.

The Role of Guidance and Special Training

What the normative interpretation of the test results cannot reveal in the case of underprivileged individuals is how much their status might change if their environmental opportunities and incentives for learning or the acquisition of skills were to be improved significantly. In the case of the Boys' Club boys mentioned above, estimates of academic growth potential are made on the basis of knowledge of the educational and cultural limitations of the boys' home and neighborhood environment, observational appraisals of the boys' behavior in club activities, and knowledge of the enhanced educational and motivational opportunities that can be offered to the boys in selected college preparatory schools. With this information available, the normative interpretation of the boys' scores on standardized tests can be tempered with experienced judgment, and better estimates of the boys' academic potential can thus be made.

In situations where minority group members are likely to have to continue competing with others under much the same cultural handicaps that they have faced in the past, normative interpretation of their aptitude and achievement test scores will probably yield a fairly dependable basis for short term predictive purposes. When special guidance or training is offered to help such individuals overcome their handicaps, however, achievement beyond the normative expectancies may well be obtained, and predictions should be based on expectancies derived specifically from the local situation. In this connection, it should be recognized that attempts to appraise

human "potential" without defining the milieu in which it will be given an opportunity to materialize are as futile as are attempts to specify the horsepower of an engine without knowing how it will be energized.

"Culture Fair" and "Unfair" - in the Test and in Society

The fact that a test differentiates between culturally disadvantaged and advantaged groups does not necessarily mean that the test is invalid. "Culturally unfair" test may be valid predictors of culturally unfair but nevertheless highly important criteria. Educational attainment, to the degree that it reflects social inequities rather than intrinsic merit, might be considered culturally unfair. However, a test must share this bias to qualify as a valid predictor. Making a test culture-fair may decrease its bias, but may also eliminate its criterion validity. The remedy may lie in the elimination of unequal learning opportunities, which may remove the bias in the criterion as well as in the test. This becomes more a matter of social policy and amelioration rather than a psychometric problem, however.

The situation is quite different for a test that differentiates between advantaged and disadvantaged groups even more sharply than does the criterion. The extreme case would be a test that discriminates between disadvantaged and advantaged groups but does not predict the criterion. Here one may speak of a true "test bias". The test is spuriously responsive to factors that are associated with cultural disadvantage but which are irrelevant to the criterion. From a purely psychometric viewpoint, the elimination of such factors from measurement would be desirable, since criterion validity would be augmented and test bias would be decreased. Thus, if an intelligence test calls for the identification of unfamiliar objects, and if such identification are particularly difficult for disadvantaged children, and if the ability to make such identifications bears little relation to school achievement, then the deletion of such items might well be in order. If,

however, the identification of such objects is related to the criterion of school achievement, elimination or revision of the pertinent items may decrease the test's criterion validity.

While certain aptitude and ability tests may have excellent criterion validity for some purposes even the best of them are unlikely to reflect the true capacity for development of underprivileged children. For, to the extent that these tests measure factors that are related to academic success, they must tap abilities that have been molded by the cultural setting. Furthermore, the test content, the mode of communication involved in responding to test items, and the motivation needed for making the responses are intrinsically dependent upon the cultural context.

Elixir of "Culture-Fair" Tests

The elixir of the "culture-fair" or "culture-free" test has been pursued through attempts to minimize the educational loading of test content and by reducing the premium on speed of response. However, these efforts have usually resulted in tests that have low validities for academic prediction purposes and little power to uncover hidden potentialities of children who do poorly on the common run of academic aptitude and achievement tests.

In spite of their typical cultural bias, standardized tests should not be sold short as a means for making objective assessments of the traits of minority-group children. Many bright, non-conforming pupils, with backgrounds different from those of their teachers, make favorable showings on achievements tests, in contradistinction to their low classroom marks. These are very often children whose cultural handicaps are most evident in their overt social and interpersonal behavior. Without the intervention of standardized tests, many such children would be stigmatized by the adverse sub-

jective ratings of teachers who tend to reward conformist behavior of middle-class character.

III. THE VALIDITY OF TEST INTERPRETATION

The most important consideration of all is one that applies to the use of tests in general - namely, that test results should be interpreted by competently trained and knowledgeable persons wherever important issues or decisions are at stake. Here, an analogy may be drawn from medical case history information that is entered on a child's record. Certain features of this record, such as the contagious-disease history, constitute factual data that are easily understood by school staff members who have not had medical training. But other aspects of the medical record, as well as the constellation of factors that contribute to the child's general state of health, are not readily interpretable by persons outside the medical profession. Consequently, the judgment of a doctor is customarily sought when an overall evaluation of the child's physical condition is needed for important diagnostic or predictive purposes. So, too, the psychological and educational test records of children should be interpreted by competently trained professional personnel when the test results are to be used as a basis for decisions that are likely to have a major influence on the child's future.

There are several sources of error in test interpretation stemming from a lack of recognition of the special features of culturally disadvantaged groups. One of these may be called the "deviation error." By this is meant the tendency to infer maladjustment or personality difficulty from responses which are deviant from the viewpoint of a majority culture, but which may be typical of a minority group. The results of a test might accurately reflect a child's performance or quality of ideation, but still the results should be interpreted in the light of the child's particular circumstance in life

and the range of his experiences. For example, a minister's son whose test responses indicate that he sees all women as prostitutes and a prostitute's son whose test responses give the same indication may both be accurately characterized in one sense by the test. The two boys may or may not be equally disturbed, however. Clinically, a safer inference might be that the minister's son is the one who is more likely to be seriously disturbed by fantasies involving sex and women.

There is ample evidence to indicate that members of a tribe that has experienced periodic famines would be likely to give an inordinate number of food responses on the Rorschach. So too might dieting Palm Beach matrons, but their underlying anxiety patterns would be quite different from those of the tribesmen. Or, to take still another example, the verbalized self-concept of the son of an unemployed immigrant might have to be interpreted very differently from that of a similar verbalization of a boy from a comfortable, middle-class, native-American home.

A performance IQ that is high in relation to the individual's verbal IQ on the Wechsler scales may signify psychopathic tendencies but it also may signify a poverty of educational experience. Perceiving drunken males beating up women on the Thematic Apperception Test may imply a projection of idiosyncratic fantasy or wish, but it may also imply a background of rather realistic observation and experience common to many minority group children.

For children in certain situations, test responses indicating a low degree of motivation or an over-submissive self-image are realistic reflections of their life conditions. If these children were to give responses more typical of the general population, they might well be regarded as subgroup deviants. In short, whether test responses reflect secondary defenses against anxiety or are the direct result of a socialization pro-

cess has profound diagnostic import so that knowledge of the social and cultural background of the individual becomes quite significant.

What Does the Test Really Measure?

A second type of error, from the viewpoint of construct and content validity[5], might be called the "simple determinant error." The error consists in thinking of the test content as reflecting some absolute or pure trait, process, factor, or construct, irrespective of the conditions of measurement or of the population being studied. Thus, a fifth-grade achievement test may measure arithmetical knowledge in a middle-class neighborhood where most children are reading up to grade level, but the same test, with the same content, may be strongly affected by a reading comprehension factor in a lower-class school and therefore may be measuring something quite different than what appears to be indicated by the test scores. Generally the test-taking motivation present in a middle-class group allows the responses to test content to reflect the differences in intelligence, achievement, or whatever the test is designed to measure. Contrariwise, in a population where test success has much less reward-value and where degree of test-taking effort is much more variable from individual to individual, the test content may tap motivation as well as the trait purportedly being measured.

Caution and knowledge are necessary for understanding and taking into account testing conditions and test-taking behavior when test results are being interpreted for children from varying backgrounds. A child coming from a particular cultural subgroup might have very little motivation to do well in most test situations, but under certain conditions or with special kinds of materials he might have a relatively high level of motivation. As a result, considerable variability might be evident in his test results from one situation to another, and his scores might be

difficult to reconcile and interpret.

How a question is asked is undoubtedly another important factor to consider in interpreting test results. A child might be able to recognize an object, but not be able to name it. Or, he might be able to identify a geometric figure, but not be able to reproduce it. Thus, different results might be obtained in a test depending upon whether the child is asked to point to the triangle in a set of geometric figures or whether he is required to draw a triangle.

Response Sets May Affect Test Results

In attitude or personality questionnaires, response sets[6] such as the tendency to agree indiscriminately with items, or to give socially desirable responses, may contribute error variance from the viewpoint of the content or behavior it is desired to sample. To the extent that such sets discriminate between socially advantaged and disadvantaged groups, the target content area may be confounded by specific test format. Thus, a scale of authoritarianism may be found to differentiate among social classes, but if the scale is so keyed that a high score on authoritarianism is obtained from agreement with items, the social class differences may be more reflective of an agreement set rather than an authoritarian tendency. If authoritarian content is logically distinct from agreement content, these two sources of test variance should be kept distinct either through statistical control, by a change in the item format, or by having more than one approach to measurement of the trait in question.

From the standpoint of content validity, there is a third type of error. This may be termed the "incompleteness of content coverage" error. This refers to a circumscribed sampling of the content areas in a particular domain. In the area of intelligence, for instance, Guilford (1954)

has identified many factors besides the "primary mental abilities" of Thurstone and certainly more than is implied in the unitary concept of intelligence reflected by a single IQ score. As Dreger and Miller (1960) point out, differences in intellectual functioning among various groups cannot be clearly defined or understood until all components of a particular content area have been systematically measured.

Familiarity with the cultural and social background of minority-group children not only helps to avoid under-evaluating the test performance of some children, but also helps to prevent over-evaluating the performance of others. For example, children who have been trained in certain religious observances involving particular vocabularies and objects, or those who have been encouraged to develop particular skills because of their cultural orientations, might conceivably score "spuriously" high on some tests or on particular items. In other words, any special overlap between the subgroup value-system of the child and the performances tapped by the test is likely to be an important determinant of the outcome of the test.

Failure Barriers May Be Encountered

Failure-inducing barriers are often set up for the minority-group child in a testing situation by requiring him to solve problems with unfamiliar tools, or by asking him to use tools in a manner that is too advanced for him. To draw an analogy, if a medical student were handed a scalpel to lance a wound, and if the student were to do the lancing properly but were to fail to sterilize the instrument first, how should he be scored for his accomplishment? If he had never heard of sterilization, should his skillful performance with the instrument nevertheless be given a "zero" score? Similarly, if a child from a disadvantaged social group shows a considerable degree of verbal facility in oral communication with his peers but does very poorly on tests that stress academic vocabulary, can he justifiably be ranked

low in verbal aptitude?

In a broad sense, most intelligence test items involve familiarity with language and symbol systems, although experience in symbol manipulation and communication skills varies enormously from one social group to another. One might reasonably expect that a child living in a community that minimizes language skills - or, as depicted by Bernstein (1960), a community that uses a language form that is highly concrete rather than abstract - will earn a score that has a meaning very different from that of the score of a child in the opposite kind of social and cultural circumstances. It is important, therefore, to interpret test results in relation to the range of situations and behaviors imposed by the environments of specific minority groups.

Some Suggested Remedies

While this analysis of the problems involved in the use and interpretation of tests for minority group children may lead to considerable uneasiness and skepticism about the value of the results for such children, it also points up potential ways of improving the situation. For example, one of these ways might consist of measuring separate skills first, gradually building up to more and more complex items and tests which require the exercise of more than one basic skill at a time. With enough effort and ingenuity, a sizable universe of items might be developed by this procedure. Further, we should depart from too narrow a conception of the purpose and function of testing. We should re-emphasize the concept of the test as an integral component of teaching and training whereby a floor of communication and understanding is established and learning capabilities are measured in repeated and cyclical fashion.

Finally, we should think in terms of making more use of everyday behavior as evidence of the coping abilities and competence of children who

do not come from the cultural mainstream. Conventional tests may be fair predictors of academic success in a narrow sense, but when children are being selected for special aid programs, or when academic prediction is not the primary concern, other kinds of behavioral evidence are commonly needed to modulate the results and implications of standardized tests.

CONCLUSION

Tests are among the most important evaluative and prognostic tools that educators have at their disposal. How unfortunate, then, that these tools are often used so routinely and mechanically that some educators have stopped thinking about their limitations and their benefits. Since the minority group child is so often handicapped in many ways his test scores may have meanings different from those of non-minority children, even when they are numerically the same. The task of the conscientious educator is to ponder what lies behind the test scores. Rather than accepting test scores as indicating fixed levels of either performance or potential, educators should plan remedial activities which will free the child from as many of his handicaps as possible. Good schools will employ well qualified pers to use good tests as one means of accomplishing this task.

In testing the minority group child it is sometimes appropriate to compare his performance with that of advantaged children to determine the magnitude of the deprivation to be overcome. At other times it is appropriate to compare his test performance with that of other disadvantaged children - to determine his relative deprivation in comparison with others who have also been denied good homes, good neighborhoods, good diets, good schools and good teachers. In most instances it is especially appropriate to compare the child's test performance with his previous test performance. Utilizing the individual child as his own control and using the test norms

principally as "bench marks", we are best able to gauge the success of our efforts to move the minority group child forward on the long, hard road of overcoming the deficiencies which have been forced upon him. Many comparisons depend upon tests, but they also depend upon our intelligence, our good will, and our sense of responsibility to make the proper comparison at the proper time and to undertake proper remedial and compensatory action as a result. The misuse of tests with minority group children is a serious breach of professional ethics. Their proper use is a sign of professional and personal maturity.

FOOTNOTES

1. See, for example, Katz (1958), Froehlich and Hoyt (1959), Anastasi (1961), Cronbach (1960), Thorndike and Hagen (1961).

2. See Anastasi (1958) and Tyler (1956).

3. For discussion of standard error of measurement, see Thorndike and Hagen (1961), Guilford (1954), Lindquist (1950).

4. Information about this program is obtainable from The Boys Club of New York, 287 East 10th Street, New York, N. Y.

5. For a discussion of various types of test validity, see Anastasi (1961), Cronbach (1960), Guilford (1954), Thorndike and Hagen (1961), Lindquist (1950).

6. For a discussion of this and related concepts, see Anastasi (1961), Cronbach (1960).

REFERENCES

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION AND NATIONAL COMMITTEE ON MEASUREMENTS USED IN EDUCATION. Technical recommendations for achievement tests. Washington, D.C., National Educational Association, 1955.

AMERICAN PSYCHOLOGICAL ASSOCIATION. Technical recommendations for psychological tests and diagnostic techniques. Psychol. Bull., 1954, 51, No. 2.

ANASTASI, A. Differential psychology. (3rd Ed.) New York: Macmillan, 1958.

ANASTASI, A. Psychological testing. (2nd Ed.) New York: Macmillan, 1961.

AULD, F., JR. Influence of social class on personality test responses. Psychol Bull., 1952, 49, 318-332.

BERNSTEIN, B. Aspects of language and learning in the genesis of the social process. J. child Psychol. Psychiat., 1961, 1, 313-324.

BERNSTEIN, B. Language and social class. Brit. J. Sociol., 1960, 11, 271-276.

CLIFFORD, P.I. and FISHMAN, J.A. What can mass testing programs do for-and-to-the pursuit of excellence in American Education? Harvard Educ. Rev., in press.

CHRONBACH, L. Essentials of psychological testing. (2nd Ed.) New York: Harper, 1960.

DEHAAM, R. and KOUGH, J. Teacher's guidance handbook: Identifying students with special needs (Vol. I. Secondary School Edition). Chicago: Science Research Associates, 1956.

DEUTSCH, M. Minority group and class status as related to social and personality factors in scholastic achievement. (Monograph No. 2) Ithaca, New York: The Society for Applied Anthropology, 1960.

DEUTSCH, M. The disadvantaged child and the learning process: some social, psychological, and developmental considerations. In H. Passow (Ed.), Education in depressed areas. New York: Teachers College Press, 1963.

DEUTSCH, M. and BROWN, B. Some data on social influences in Negro-white intelligence differences. J. Social Issues, in press.

DREGER, R. and MILLER, K. Comparative psychological studies of Negroes and Whites in the United States. Psychol. Bull., 1960, 57, 361-402.

EELS, K., ET AL. Intelligence and cultural differences. Chicago: University of Chicago Press, 1951.

FROEHLICH, C. and HOYT, K. Guidance testing (3rd Ed.). Chicago: Science Research Associates, 1959.

GUILFORD, J. Psychometric methods (2nd Ed.). New York: McGraw-Hill, 1954.

HAGGARD, E. Social status and intelligence: an experimental study of certain cultural determinants of measured intelligence. Genet. Psychol. Monogr., 1954, 49, 141-186.

KLINEBERG, O. Race differences. New York: Harper, 1935.

KATZ, M. Selecting an achievement test: Principles and procedures. Princeton: Educational Testing Service, 1958.

LINDQUIST, E. (Ed.) Educational Measurement. Washington: American Council of Education, 1950.

MASLAND, R., SARASON, S. and GLADWIN, T. Mental subnormality. New York: Basic Books, 1958.

SARASON, S. et al. Anxiety in elementary school children. New York: Wiley, 1960.

THORNDIKE, R., and HAGEN, E. Measurement and evaluation in psychology and education. (2nd Ed.) New York: Wiley, 1961.

TYLER, L. The psychology of individual differences (2nd Ed.). New York: Appleton, Century, Crofts, 1956.